Centers for Disease Control and Prevention
Epidemiology Program Office
Case Studies in Applied Epidemiology
No. 731-703

Cigarette Smoking and Lung Cancer

Instructor's Guide

Learning Objectives

After completing this case study, the participant should be able to:

- ☐ Discuss the elements of study design, and the advantages and disadvantages of case-control versus prospective cohort studies;
- ☐ Discuss some of the biases that might have affected these studies;
- ☐ Calculate a rate ratio, rate difference, odds ratio, and attributable risk percent;
- ☐ Interpret each measure and describe each measure's main use; and
- ☐ Review the criteria for causation.

This case study is based on the classic studies by Doll and Hill that demonstrated a relationship between smoking and lung cancer. Two case studies were developed by Clark Heath, Godfrey Oakley, David Erickson, and Howard Ory in 1973. The two case studies were combined into one and substantially revised and updated by Nancy Binkin and Richard Dicker in 1990. Current version updated by Richard Dicker with input from Julie Magri and the 2003 EIS Summer Course instructors.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service

A causal relationship between cigarette smoking and lung cancer was first suspected in the 1920s on the basis of clinical observations. To test this apparent association, numerous epidemiologic studies were undertaken between 1930 and 1960. Two studies were conducted by Richard Doll and Austin Bradford Hill in Great Britain. The first was a case-control study begun in 1947 comparing the smoking habits of lung cancer patients with the smoking habits of other patients. The second was a cohort study begun in 1951 recording causes of death among British physicians in relation to smoking habits. This case study deals first with the case-control study, then with the cohort study.

Data for the case-control study were obtained from hospitalized patients in London and vicinity over a 4-year period (April 1948 - February 1952). Initially, 20 hospitals, and later more, were asked to notify the investigators of all patients admitted with a new diagnosis of lung cancer. These patients were then interviewed concerning smoking habits, as were controls selected from patients with other disorders (primarily non-malignant) who were hospitalized in the same hospitals at the same time.

Data for the cohort study were obtained from the population of all physicians listed in the *British Medical Register* who resided in England and Wales as of October 1951. Information about present and past smoking habits was obtained by questionnaire. Information about lung cancer came from death certificates and other mortality data recorded during ensuing years.

Question 1: What makes the first study a case-control study?

Answer 1
In a case-control study, people diagnosed as having a disease (in this case lung cancer) are compared with others who do not have the disease (controls). The purpose is to determine if the two groups differ in the proportion of persons who had been exposed to a specific factor (in this instance, cigarette smoking).

Question 2: What makes the second study a cohort study?

Answer 2
In a cohort study, either an entire population is enrolled and participants are categorized by exposure, or participants are enrolled on the basis of their exposure status. Occurrence of disease in the different exposure groups is then ascertained, either by following participants over time (prospective study) or by ascertaining disease that has already occurred (retrospective study).

The remainder of Part I deals with the case-control study.

Question 3: Why might hospitals have been chosen as the setting for this study?

Answer 3
Reasons might have included:
- high likelihood of finding cases, ease of finding cases
- accurate diagnosis
- access to medical records, relatively complete medical records
- captive audience for study (likely to cooperate)
- convenient and plentiful source for controls

Question 4: What other sources of cases and controls might have been used?

Answer 4
CASES: cancer registries, death certificates, pathology labs, insurance files, doctors' offices, occupational records
CONTROLS: neighbors, friends / acquaintances, other patients of the same doctors, population-based

Question 5: What are the advantages of selecting controls from the same hospitals as cases?

Answer 5
- convenient
- likely to come from the same population as the cases ("If they had developed lung cancer, they would have been captured as cases.") In effect, controls for socioeconomic status, place of residence, access to care, diagnostic practices.
- temporal match
- comparable records
- likely to be cooperative (hospitalized patients = captive audience with time on their hands)
- equally heightened recall?

Question 6: How representative of all persons with lung cancer are hospitalized patients with lung cancer?

Answer 6
INSTRUCTOR'S NOTE: Do not spend too much time on Questions 6 and 7.

Fairly representative, at least at the time of the study, since most persons who developed lung cancer were hospitalized at some point. However, hospitalized patients with lung cancer may be sicker or later in disease, may have more complications or other diseases, or conversely, may be less sick (survivors).

Question 7: How representative of the general population without lung cancer are hospitalized patients without lung cancer?

Answer 7
In general, hospitalized patients are not very representative of the general population. They are more likely to smoke cigarettes and drink alcohol than the general population.

Question 8: How may these representativeness issues affect interpretation of the study's results?

Answer 8
The purpose of a control group in a case-control study is to provide the prevalence of exposure in the population from which the cases are drawn. Hospitalized controls may be in the hospital for other smoking-related diagnoses; the prevalence of smoking in a hospitalized population is greater than that found in the general population. The net effect of a higher prevalence of smokers among the controls is that the true risk of lung cancer associated with smoking will be underestimated. The resulting bias can be classified as a form of selection bias.

Over 1,700 patients with lung cancer, all under age 75, were eligible for the case-control study. About 15% of these persons were not interviewed because of death, discharge, severity of illness, or inability to speak English. An additional group of patients were interviewed but later excluded when initial lung cancer diagnosis proved mistaken. The final study group included 1,465 cases (1,357 males and 108 females).

The following table shows the relationship between cigarette smoking and lung cancer among male cases and controls.

Table 1. Smoking status before onset of the present illness, lung cancer cases and matched controls with other diseases, Great Britain, 1948-1952.

	Cases	Controls
Cigarette smoker	1,350	1,296
Non-smoker	7	61
Total	1,357	1,357

Question 9: From this table, calculate the proportion of cases and controls who smoked.

Proportion smoked, cases:

Proportion smoked, controls:

Answer 9

Proportion smoked, cases: 1,350 / 1,357 = 99.5%

Proportion smoked, controls: 1,296 / 1,357 = 95.5%

Question 10: What do you infer from these proportions?

Answer 10

Although cases have a slightly higher proportion of smokers than controls, the proportions are remarkably close. Note the overall prevalence of smoking (over 95%)!

Question 11a: Calculate the odds of smoking among the cases.

Answer 11a

Odds is a statistical (and gambling!) term calculated as the probability of something happening (e.g., being exposed, winning the race) divided by the probability of that something NOT happening. Since the probability of something not happening is 1 minus the probability of it happening, the formula for odds is:

Odds = probability / (1 - probability) = proportion / (1 - proportion)

The probabilities (proportions) of smoking among cases and controls were calculated in Question 9. The odds could be calculated as:

Odds of smoking, cases:
(1350 / 1357) / (7 / 1357) = 1350 / 7 = 192.9 : 1

Note that the total number of events or observations is in both the numerator and denominator, and therefore cancels out. The odds formula can therefore be simplified to:

Odds = # yes / # no = # wins / # losses = # exposed / # unexposed

Because the odds is a ratio of 2 probabilities, one could say that, "Case-patients were 192 times as likely to smoke as to not smoke."

Question 11b: Calculate the odds of smoking among the controls.

Answer 11b

Odds of smoking, controls:
(1296 / 1357) / (61 / 1357) = 1296 / 61 = 21.2 : 1

Question 12: Calculate the ratio of these odds. How does this compare with the cross-product ratio?

Answer 12

Ratio of odds = (a/c) / (b/d) = (1350 / 7) / (1296 / 61) = 192.9 / 21.2 = 9.1

Cross-product ratio = (a x d / b x c) = (1350 x 61) / (1296 x 7) = 9.1, i.e., algebraically identical

FYI, Mantel-Haenszel X^2 = 44.0, Cornfield 95% CI = (4.0, 21.8)

Because the odds ratio is a ratio of 2 odds, not probabilities, one CANNOT use "likely," which implies probability. One has to say something like "The odds of smoking among case-patients was 9.1 times as high as the odds of smoking among controls."

Question 13: What do you infer from the odds ratio about the relationship between smoking and lung cancer?

Answer 13

The strict wording of the odds ratio in English is provided in the previous question. Question 13 asks about inference. In a way, this is comparable to the difference between the Results section of a manuscript and its Discussion section. So strictly speaking, the odds of being a smoker are 9.1 times as high in lung cancer cases than among non-cases. In this instance, lung cancer is a rare disease and the odds ratio is a good approximator of the rate ratio. Assuming that the study is not biased, one can infer that the risk of lung cancer appears to be about 9 times as high in cigarette smokers than in non-smokers.

Table 2 shows the frequency distribution of male cases and controls by average number of cigarettes smoked per day.

Table 2. Most recent amount of cigarettes smoked daily before onset of the present illness, lung cancer cases and matched controls with other diseases, Great Britain, 1948-1952.

Daily number of cigarettes	# Cases	# Controls	Odds Ratio
0	7	61	referent
1-14	565	706	____
15-24	445	408	____
25+	340	182	____
All smokers	1,350	1,296	____
Total	1,357	1,357	

Question 14: Compute the odds ratio by category of daily cigarette consumption, comparing each smoking category to nonsmokers.

Answer 14

Instructor's Note: Split class into 4 groups. Have each group compute the odds ratio for one dose category (the fourth group do "All Smokers.") Suggest that they start by drawing a 2-by-2 table.

1-14 cigarettes, OR = (565 x 61) / (706 x 7) = **7.0**
15-24 cigarettes, OR = (445 x 61) / (408 x 7) = **9.5**
25+ cigarettes, OR = (340 x 61) / (182 x 7) = **16.3**
All smokers, OR = (1350 x 61) / (1296 x 7) = **9.1**

Question 15: Interpret these results.

Answer 15
Values of the odds ratio rise steadily, consistent with a dose-response relationship between the daily number of cigarettes smoked and the strength of the association.

Although the study demonstrates a clear association between smoking and lung cancer, cause-and-effect is not the only explanation.

Question 16: What are the other possible explanations for the apparent association?

Answer 16
Explanations other than a true association are:
- Chance (although the statistical tests indicate that chance is an unlikely explanation)
- Selection bias
- Information bias
- Confounding
- Investigator error

An example of a likely **selection bias** in this study is that the controls were chosen from among hospitalized patients, who are more likely to be smokers than the general population. (The effect of the bias, however, would be to underestimate, rather than overestimate the risks associated with smoking.)

Information bias could have occurred if lung cancer cases were more likely to accurately recall their smoking history than the controls. Such a bias is not highly likely in this instance since the hypothesis regarding an association between smoking and lung cancer was not widely known and also because controls were other hospitalized patients who were probably as likely as the cases to be introspective about previous exposures or events.

Age might be a potential **confounder** in this study. To be a confounder, a factor must be associated with, but not a consequence of, an exposure, and independent of its association with the exposure, must also be associated with the outcome. If lung cancer is more likely to occur among older people and being older is associated with an increased likelihood of being a smoker, then the observed association between smoking and lung cancer might simply reflect the association between age and lung cancer.

Investigator error due to transcription or data entry errors, or inappropriate analyses is always possible, but the authors have a well-deserved reputation for their careful methods and analysis.

The next section of this case study deals with the cohort study.

Data for the cohort study were obtained from the population of all physicians listed in the British Medical Register who resided in England and Wales as of October 1951. Questionnaires were mailed in October 1951, to 59,600 physicians. The questionnaire asked the physicians to classify themselves into one of three categories: 1) current smoker, 2) ex-smoker, or 3) nonsmoker. Smokers and ex-smokers were asked the amount they smoked, their method of smoking, the age they started to smoke, and, if they had stopped smoking, how long it had been since they last smoked. Nonsmokers were defined as persons who had never consistently smoked as much as one cigarette a day for as long as one year.

Usable responses to the questionnaire were received from 40,637 (68%) physicians, of whom 34,445 were males and 6,192 were females.

Question 17: How might the response rate of 68% affect the study's results?

Answer 17
As a general rule of thumb, we like to see response rates of 80% or better in epidemiologic studies. Realistically, a 68% response rate is good for a mail study. If participation in a prospective cohort study is not related to both exposure and disease status, then a suboptimal response rate will only decrease the power of the study and will not bias the measure of association. If participation in a prospective cohort study is related to both exposure and disease status, then selection bias may be a problem. Therefore, if possible, you should characterize the nonrespondents as best you can and determine whether the respondents differ on important factors.

The next section of this case study is limited to the analysis of male physician respondents, 35 years of age or older.

The occurrence of lung cancer in physicians responding to the questionnaire was documented over a 10-year period (November 1951 through October 1961) from death certificates filed with the Registrar General of the United Kingdom and from lists of physician deaths provided by the British Medical Association. All certificates indicating that the decedent was a physician were abstracted. For each death attributed to lung cancer, medical records were reviewed to confirm the diagnosis.

Diagnoses of lung cancer were based on the best evidence available; about 70% were from biopsy, autopsy, or sputum cytology (combined with bronchoscopy or X-ray evidence); 29% were from cytology, bronchoscopy, or X-ray alone; and only 1% were from just case history, physical examination, or death certificate.

Of 4,597 deaths in the cohort over the 10-year period, 157 were reported to have been caused by lung cancer; in 4 of the 157 cases this diagnosis could not be documented, leaving 153 confirmed deaths from lung cancer.

The following table shows numbers of lung cancer deaths by daily number of cigarettes smoked at the time of the 1951 questionnaire (for male physicians who were nonsmokers and current smokers only). Person-years of observation ("person-years at risk") are given for each smoking category. The number of cigarettes smoked was available for 136 of the persons who died from lung cancer.

Table 3. Number and rate (per 1,000 person-years) of lung cancer deaths by number of cigarettes smoked per day, Doll and Hill physician cohort study, Great Britain, 1951-1961.

Daily number of cigarettes smoked	Deaths from lung cancer	Person-years at risk	Mortality rate per 1000 person-years	Rate Ratio	Rate difference per 1000 person-years
0	3	42,800	0.07	referent	referent
1-14	22	38,600	0.57	8.1	0.50
15-24	54	38,900	1.39	19.8	1.32
25+	57	25,100	2.27	32.4	2.20
All smokers	133	102,600	1.30	18.6	1.23
Total	136	145,400	0.94		

Question 18: Compute lung cancer mortality rates, rate ratios, and rate differences for each smoking category. What do each of these measures mean?

Answer 18
See completed table above.

INSTRUCTORS NOTES: YOU MAY WANT TO REVIEW THE CONCEPT OF PERSON-YEAR WITH THE STUDENTS (Can be interpreted as "per xxx persons per year").

- Point out that it is possible to directly calculate rates here, including rates for each exposure category.
- Data show lung cancer mortality rates increase with amount smoked.

Mortality rate reflects the mortality experience of the population over time. It is a true rate in the sense that it conveys the velocity at which the population is dying, e.g., 0.94 deaths per 1,000 persons per year, or almost 1 death per 1,000 persons per year.

The **rate ratio** is a measure of strength of association showing proportionate increase in rate of disease with increasing exposure. "The mortality rate is 18.4 times as high among smokers than among non-smokers (!)"

The **rate difference** is the difference between the rates in the exposed and unexposed groups and provides information about the excess risk of a disease attributable to the exposure. "Overall, smoking causes 1.23 excess deaths per 1,000 persons per year." Excess deaths increase from 0.5 to 2.2 per 1,000 person-years as the quantity of daily cigarettes smoked increases.

Question 19: What proportion of lung cancer deaths among all smokers can be attributed to smoking? What is this proportion called?

Answer 19

This proportion is called (by different epidemiologists):

- *the attributable risk percent* (Hennekens; Greenberg; most people at CDC)
- *attributable proportion* for the exposed population (Rothman)
- *attributable fraction for the exposed* (Kelsey-Thompson-Evans)
- *etiologic fraction in the exposed* (Miettinen; Kleinbaum-Kupper-Morgenstern)
- *excess fraction* (Greenland)
- *proportion attributable risk* (Gordis)

We will use the term **attributable risk percent** (AR%) in this Instructor's Guide. It is usually calculated in one of two ways:

$$AR\% = (Incidence_{exposed} - Incidence_{unexposed}) / Incidence_{exposed}$$

$$AR\% = (RR - 1) / RR$$

Using the first formula, AR% = (1.30-0.07)/1.30 or 0.95. Thus, most, but not all of the lung cancer deaths among smokers are due to smoking.

Question 20: If no one had smoked, how many deaths from lung cancer would have been averted?

Answer 20

Overall 95% of lung cancer deaths among smokers are attributable to smoking. Therefore, if none of the smokers had smoked, 95% of 133 deaths, or 126 deaths, would have been averted.

Alternatively, under the null hypothesis that smokers should have the same mortality rate as non-smokers (0.07), the expected number of deaths among smokers would be (0.07/1,000) x 102,600 = 7 deaths. Subtracting 7 expected or baseline deaths from the 136 observed deaths among smokers yields 126 excess deaths that could be averted.

The cohort study also provided mortality rates for cardiovascular disease among smokers and nonsmokers. The following table presents lung cancer mortality data and comparable cardiovascular disease mortality data.

Table 4. Mortality rates (per 1,000 person-years), rate ratios, and excess deaths from lung cancer and cardiovascular disease by smoking status, Doll and Hill physician cohort study, Great Britain, 1951-1961.

	Mortality rate per 1,000 person-years					
	Smokers	Non-smokers	All	Rate ratio	Excess deaths per 1,000 person-years	Attributable risk percent among smokers
Lung cancer	1.30	0.07	0.94	18.5	1.23	95%
Cardiovascular disease	9.51	7.32	8.87	1.3	2.19	23%

Question 21: Which cause of death has a stronger association with smoking? Why?

Answer 21
The rate ratio is the primary measure of association. Thus there is a much stronger association between smoking and lung cancer mortality than between smoking and cardiovascular mortality as indicated by a 14-fold greater rate ratio for smokers (18.5 versus 1.3.)

In calculating the **attributable risk percent**, the excess lung cancer deaths attributable to smoking is expressed as a percentage of all lung cancer mortality among all smokers. The attributable risk percent of 95% for smoking may be interpreted as the proportion of lung cancer deaths among smokers that could have been prevented if they had not smoked.

A similar measure, the **population attributable risk percent** expresses the excess lung cancer deaths attributable to smoking as a percentage of all lung cancer mortality among the entire population. From a prevention perspective, the population attributable risk percent for a given exposure can be interpreted as the proportion of cases in the entire population that would be prevented if the exposure had not occurred. The population attributable risk percent is often used in assessing the cost-effectiveness and cost-benefit of community-based intervention programs.

One formula for the population attributable risk percent is:

PAR% = (Incidence in entire population – Incidence in unexposed) / Incidence in entire population

Question 22: Calculate the population attributable risk percent for lung cancer mortality and for cardiovascular disease mortality. How do they compare? How do they differ from the attributable risk percent?

Answer 22
Population attributable risk percent is the percentage of disease/death in the population that is attributable to an exposure. There are many formulas that can be used to calculate this value:

PAR% = (Incidence in entire pop. – Incidence in unexposed) / Incidence in entire pop.

= (% exposed among cases) × ($Incidence_{exposed}$ – $Incidence_{unexposed}$) / $Incidence_{exposed}$

= (% exposed among cases) × attributable risk percent

= (% exposed among cases) × (RR – 1) / RR

= P(RR – 1) / [P(RR – 1) + 1] where P = % exposed among entire population

Using the first formula,
for lung cancer, PAR% = (0.94 - 0.07) / 0.94 = 0.925 or 92.5%;
for cardiovascular disease, PAR% = (8.87 - 7.32) / 8.87 = 0.174 or 17.4%.
In words, 92.5% of all lung cancer deaths in the population and 17.4% of all cardiovascular disease deaths in the population are attributable to smoking.

By contrast, the attributable risk percent (see Question 19) is the percentage of the disease among the exposed group that is attributable to the exposure. In this case, 95% of lung cancer deaths among smokers can be attributed to the fact that they smoke, while 23% of cardiovascular deaths among smokers can be attributed to their smoking. Note that the values for attributable risk percent and population attributable risk percent are similar because the prevalence of smoking in the study population is so high. If the prevalence of smoking were lower, the population attributable risk percent would be lower, while the attributable risk percent is unaffected by prevalence.

Question 23: How many lung cancer deaths per 1,000 persons per year are attributable to smoking among the entire population? How many cardiovascular disease deaths?

Answer 23

One way of addressing this question is to determine the rate of each disease in the population attributable to the exposure, smoking, by multiplying the population attributable risk percent times the rate of disease in the population.

For lung cancer, 0.94 / 1,000 PY × 0.925 = 0.87 lung cancer deaths per 1,000 pop. per year

For CVD, 8.87 / 1,000 PY × 0.174 = 1.54 cardiovascular deaths per 1,000 pop. per year

The number of smoking-related deaths per 1000 person-years is greater for cardiovascular disease than for lung cancer even though the rate ratio is considerably lower. Thus, if no one smoked, more cardiovascular deaths would be prevented than lung cancer deaths.

INSTRUCTOR'S NOTE:

Preventable fraction is an estimate of what might be achieved by implementing a public health intervention program in a community setting. **Prevented fraction** is a measure of what actually has been achieved after the intervention has been implemented. Both are calculated by the same formula:

PF = (Incidence in unexposed – Incidence in entire pop.) / Incidence in unexposed
= (% exposed in population) x (1 - RR)

where exposed refers to exposure to the intervention. To calculate preventable fraction, use the estimated % exposed in population. To calculate prevented fraction, use the observed % exposed in population.

PF in the exposed is the comparable calculation limited to the exposed group. Vaccine efficacy is an example of PF in the exposed.

PF_{exp} = (Incidence in unexposed – Incidence in exposed) / Incidence in unexposed
= 1 - RR

The following table shows the relationship between smoking and lung cancer mortality in terms of the effects of stopping smoking.

Table 5. Number and rate (per 1,000 person-years) of lung cancer deaths for current smokers and ex-smokers by years since quitting, Doll and Hill physician cohort study, Great Britain, 1951-1961.

Cigarette smoking status	Lung cancer deaths	Rate per 1000 person-years	Rate Ratio
Current smokers	133	1.30	18.5
For ex-smokers, years since quitting:			
<5 years	5	0.67	9.6
5-9 years	7	0.49	7.0
10-19 years	3	0.18	2.6
20+ years	2	0.19	2.7
Nonsmokers	3	0.07	1.0 (ref)

Question 24: What do these data imply for the practice of public health and preventive medicine?

Answer 24
The lowest rate is seen among those who never smoked. However, although the lung cancer mortality rate decreases with time since last smoked, even after 20 years of abstinence the rate is nearly three times greater than for never smokers.

Hence, smoking cessation efforts are worthwhile from a public health point of view, but smoking prevention efforts would be most valuable.

As noted at the beginning of this case study, Doll and Hill began their case-control study in 1947. They began their cohort study in 1951.

The odds ratios and rate ratios from the two studies by numbers of cigarettes smoked are given in the table below.

Table 6. Comparison of measures of association from Doll and Hill's 1948-1952 case-control study and Doll and Hill's 1951-1961 physician cohort study, by number of cigarettes smoked daily, Great Britain.

Daily number of Cigarettes smoked	Rate ratio from cohort study	Odds ratio from case-control study
0	1.0 (ref)	1.0 (ref)
1-14	8.1	7.0
15-24	19.8	9.5
25+	32.4	16.3
All smokers	18.5	9.1

Question 25: Compare the results of the two studies. Comment on the similarities and differences in the computed measures of association.

Answer 25

The odds ratios in the case-control study consistently underestimate the rate ratios, probably because of the use of hospital patients as controls (hospitalized controls with other diseases were very likely to be smokers).

However, overall, the two studies provide very consistent results, including evidence of a dose-response effect in both.

Instructor's Note: Mathematically, the odds ratio always overestimates the risk ratio when both are based on the same population. You can see for yourself by calculating the odds ratio and risk ratio from the same cohort study data in a 2-by-2 table. The high the prevalence of illness in the population, the greater the overestimation. Shown below is the 2-by-2 table for vanilla ice cream and illness (overall prevalence very high at 61.3%) from the outbreak of gastroenteritis in Oswego, N.Y.

		Ill	Well	Total	Attack rate
Ate vanilla ice cream?	Yes	43	11	54	79.6%
	No	3	18	21	14.3%

Odds ratio = 43 x 18 / (3 x 11) = 23.45
Risk ratio = 79.6 / 14.3 = 5.6

So, the fact that the odds ratios underestimate the rate ratios in the smoking and lung cancer studies probably reflects how much the hospital controls were unrepresentative of the general population in terms of smoking prevalence. In spite of this substantial selection bias that weighed against finding a significant association, the association is so large that the case-control study still found one.

Question 26: What are the advantages and disadvantages of case-control vs. cohort studies?

Answer 26

	Case-control	Cohort
Sample size	**small**	**large**
Costs	**less**	**more**
Study time	**short**	**long**
Rare disease	**advantage**	**disadvantage**
Rare exposure	**disadvantage**	**advantage**
Multiple exposures	**advantage**	**disadvantage**
Multiple outcomes	**disadvantage**	**advantage**
Progression, spectrum of illness	**disadvantage**	**advantage**
Disease rates	**cannot measure**	**advantage**
Recall bias	**potential problem**	**less problem**
Loss to follow-up	**advantage**	**potential problem**
Selection bias	**potential problem**	**less problem**

Question 27: Which type of study (cohort or case-control) would you have done first? Why? Why do a second study? Why do the other type of study?

Answer 27

First, a case-control study is quicker and easier. If the case-control study provides results which warrant further investigation, then it is appropriate to do a second study to confirm the findings. The cohort study, which is more difficult and expensive to mount and is slower to yield results, provides confirmation, better assessment of natural progression from exposure to disease, allows calculation of disease rates, and, depending on choice of study subjects, may be more generalizable.

Question 28: Which of the following criteria for causality are met by the evidence presented from these two studies?

Answer 28

	YES	NO
Strong association	**X**	
Consistency among studies	**X**	
Exposure precedes disease	**X**	
Dose-response effect	**X**	
Biologic plausibility		**X**

Instructor's Note: The X in the NO column for biologic plausibility does not mean that the association between smoking and lung cancer is not biologically plausible, just that the biologic plausibility evidence has not been presented in this case study.

REFERENCES

1. Doll R, Hill AB. Smoking and carcinoma of the lung. *Brit Med J* 1950; 2:739-748.

2. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *Brit Med J* 1952; 2:1271-1286.

3. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits. *Brit Med J* 1954; 1:1451-1455.

4. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking. *Brit Med J* 1956; 2:1071-1081.

5. Doll R, Hill AB. Mortality in relation to smoking: 10 years' observation of British doctors. *Brit Med J* 1964; 1:1399-1410, 1460-1467.

6. U. S. Public Health Service. Smoking and health. Report of the Advisory Committee to the Surgeon General of the Public Health Service. US Department of Health, Education, and Welfare, PHS, CDC. PHS Publication No. 1103, 1964.

7. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.

8. Levy RA, Marimont RB. Lies, damned lies, and 400,000 smoking-related deaths. *Regulation* 1998; 21-29.